

UNDERSTANDING MADNESS?

Simon J. Evnine

Irrationality, usually in the form of self-deception or weakness of the will, is an issue which has been much debated in philosophy in recent years. The kinds of cases discussed, however, are almost always of a limited nature, isolated episodes in the lives of otherwise normal, rational people. The question of madness, of a large-scale breakdown in rationality, by contrast, has received little or no attention. The apparent problem posed by madness is one of understanding. In the third part of this paper, I will develop a more technical notion of understanding, but until then, we can use the concept quite informally and intuitively. The insane seem to have beliefs and desires which, for various reasons, we find it hard to see the justification for, or the point of. How can someone really believe of himself that he is a gourd, or that he is made of glass? How can someone really have the desire to eat up the world? There may also be phenomena which are even more difficult to understand, such as belief in an apparent contradiction, or non-transitive patterns of preference.

In this paper, I propose to look at two suggestions which have been made – by philosophically inclined psychiatrists – about how to understand the psychotically insane. I will argue that the methods proposed are unsatisfactory, though for different reasons. Finally, I will say something about how a particular philosophical approach to the topic of mental content might have something useful to say about the problem of understanding madness.

Before I begin, a brief *caveat*: in what follows, I use the term “psychosis” and its cognates interchangeably with “madness” and “insanity” and their cognates. And I use none of these in any clinically defined sense. Indeed, whether or not these terms *can* be given any precise clinical sense is largely an irrelevance here since I am happy for my remarks to apply to any cases which present the difficulties for understanding which I take as the sign of madness, regardless of whether they match any proposed clinical definition.

I

The first method for understanding psychotics is particularly associated with R. D. Laing. Laing discusses the case of someone

who says of himself that he is an “unreal man”. This is not a case of someone who is “lying, joking, or equivocating in some subtle way”.¹ The man who utters the sentence “I am an unreal man” means this as a straightforwardly true assertion. There is apparently something quite irrational about such a person. Sincerely asserting a sentence is a way of expressing a belief. The belief which the English sentence “I am unreal” expresses is the belief that its speaker is unreal. But this is where the irrationality lies. We cannot understand how someone could have the belief that he is unreal. This is why, when someone utters the sentence “I am unreal” we take him to be mad.

Of course, we need not suppose that the psychotic simply asserts one maverick sentence. He may assert many other sentences, and perform many actions, which are associated with his adherence to the sentence in question. And it may be that, given that he has the belief expressed by that sentence, his other beliefs and actions are, in an uncontroversial sense, quite rational. Given the psychotic’s belief that he is unreal, together, say, with the belief that he exists, we can quite well understand his drawing the conclusion that something unreal exists even if this latter belief is also, in itself, irrational. Or, to adapt an example from Foucault, if someone thinking that he is dead, and that the dead don’t eat, doesn’t eat even though he starves himself to death by not eating, we can see how his action of not eating is, although in itself irrational, at least rational in relation to his other beliefs. The trouble, in both these cases, is with the original belief that he is unreal, or dead.

Laing thinks that a person who utters such a sentence is trying to tell us something. It would be a serious mistake to dismiss the utterance, or belief, as mad. Changing the example from someone who says “I am unreal” to someone who says “I am dead”, Laing writes:

A man says he is dead but he *is* alive. But his “truth” is that he is dead. He expresses it perhaps in the only way common (i.e. communal) sense allows him. He means that he is “really” and quite “literally” dead, not merely symbolically or “in a sense” or “as it were”, and is seriously bent on communicating his truth. The price, however, to be paid for transvaluating the communal truth in this manner is to “be” mad, for the only *real* death *we* recognize is biological death.²

¹ *The Divided Self* (Harmondsworth: Penguin Books, 1964), p. 36.

² *Ibid.*, pp. 37–8.

The abundance of scare-quotes makes this a difficult paragraph to unravel, but I take Laing to mean something like this. A man confronts us and utters the words "I am dead". Our first response is to suppose that this is a plain English sentence. The belief that an assertion of this sentence expresses is the belief that the speaker is dead. The only way of being dead we, the sane, recognize, is being biologically dead and no sane person can believe he is biologically dead. So we might suppose the speaker intends his words metaphorically, or with an "as it were". But we rule this out. (Laing doesn't give any indication of the criteria for something's being meant literally.) Our conclusion is that the speaker must just be mad. He really does have the crazy belief that he is biologically dead, although it is clear he is not. Such a person is deluded and to that extent we fail to understand him.

But, Laing goes on to imply in the passage quoted, there is another possible response. Suppose a creature stepped off a spaceship and uttered the sounds which make up the sentence "I am dead". We would not automatically take this alien to have the deluded belief that he was dead. We would probably just think that, by an amazing coincidence, a sentence of the alien's language happened to sound like a sentence of ours. We would not suppose that for his sentence to be true, the speaker had to be biologically dead. To get the point of the alien's utterance, we would have to try and ascertain what he thought had to be the case for his sentence to be true. One could put this by saying that we had to find out "his truth", meaning thereby, find out the truth-conditions of his sentence.

This is what Laing thinks we should do with the psychotic. On the one hand "his 'truth' is that he is dead", i.e., for him, his sentence "I am dead" is straightforwardly true. But on the other hand, he has transvaluated the communal truth since for him, biological death is not the only real death. Putting these two things together, we can say that when he utters the sentence "I am dead" he is making a sincere assertion, expressing what he takes to be a truth, but that his sincere assertion, "his 'truth' ", is not that he is biologically dead. He is, therefore, uttering a sentence which has different truth-conditions from our sentence "I am dead". Whereas our sentence is true if and only if its speaker is biologically dead, his is true under other conditions. (Much of Laing's book is an attempt to say under what conditions the psychotic's sentence is true.) The psychotic, it turns out, is merely a speaker of a foreign language. To understand him, we must do just what we would have to do to

understand an alien from another planet or a person from an undiscovered tribe. We must decide what his sentences mean by ascribing to them truth-conditions – we must discover “his ‘truth’ ”.

This conclusion, that the psychotic is really just speaking another language, is no doubt not unwelcome to Laing. He endorses Sullivan’s view that the psychotic is, “more than anything else ‘simply human’ ”.³ The point of treating the psychotic’s utterance of “I am dead” as an utterance of a non-English sentence is precisely to show that he is much less strange than we thought. It may be that an utterance even of his own sentence “I am dead” reveals a certain strangeness about him. After all, the alien who steps off his ship and utters the sounds “I am dead” may turn out to have a crazy belief, not because his sentence expresses the belief that he is dead, but because, in Alienese, it expresses the belief that his head is a raspberry ripple. (Or perhaps it expresses this belief, but his head really is a raspberry ripple, so his having that belief is not crazy.)

Whether or not the psychotic’s utterance of “I am dead” is strange will depend on exactly what his sentence turns out to mean. Sometimes Laing writes as if it is not such a strange sentence to utter. A therapist treating a psychotic must draw “on his own psychotic possibilities, without forgoing his sanity”.⁴ At other times he admits that “the kernel of the schizophrenic’s experience of himself must remain incomprehensible to us”.⁵ But whatever strangeness there is in the psychotic, it will not be evident merely from his uttering the sentence “I am dead”, since by asserting this sentence, he is not expressing the belief that he is biologically dead, and it was on the supposition that he held this belief that we originally thought him mad.

What is odd about a grown-up person, whom everyone thought spoke English, uttering the sentence “I am dead”? For Laing, the answer must be “in itself, nothing”. There may, or may not, be something wrong with believing whatever belief he is expressing by his assertion. This is a matter for investigation, and the answer cannot be given without first finding out what he means. But merely by virtue of uttering the sentence “I am dead”, the psychotic has given no evidence of his psychosis. It is this

³ *Ibid.*, p. 34.

⁴ *Ibid.*, p. 34.

⁵ *Ibid.*, p. 38.

implication which seems to make Laing's approach to understanding psychosis deficient. It fails to locate the oddness of psychosis at the right place. Of course, Laing is not denying that in important respects the psychotic is different from many of those considered sane by common consent. His account of ontological insecurity is aimed at uncovering these differences. But it remains the case that for Laing, one of the very places where we had located the person's madness, his utterance of the sentence "I am dead", becomes, in itself, innocuous.

Laing does attempt to give some account of what is wrong with the psychotic's utterance of a sentence like "I am dead" but of course, by substituting talk of "being" mad for talk of being mad, Laing is implying that the source of the wrongness is entirely with those who hear the utterance, and not with those who make it. In the paragraph quoted at some length, he writes that to "be" mad is "the price . . . to be paid for transvaluating the communal truth". The phrase "transvaluating the communal truth" is, despite its crucial importance here, ambiguous and ultimately more rhetorical than informative. There is the suggestion that what the psychotic does is to hold true what the community holds false. But however reprehensible our practice in this instance, no-one could theorize that psychosis is simply a matter of disagreement with the community. It is true that there are some beliefs that are held by the community, and such that if anyone differs in his evaluation of the truth of these beliefs, he would be considered mad. Such beliefs might include, for instance, every conscious living person's beliefs that he is real and alive. But of course, on Laing's analysis, it is not these beliefs which the psychotic holds false, for his truth is different, and for him, death is not just a matter of biological death.

Another more interesting implication of the phrase "transvaluating the communal truth" is that the psychotic has to pay the price of "being" mad, not simply for speaking a foreign language, but for speaking a foreign language whose sentences bear too uncannily a resemblance to the language of his community. This may be what Laing means, since why else should he think that someone must pay a price for what he diagnoses as merely speaking a foreign language.

I don't think this idea is quite as ridiculous as it may first appear. Consider the members of some subversive revolutionary party, intent on undermining the society in which they live. They might develop a language whose sentences were homophonic to English sentences, but such that if utterances of them were interpreted as

utterances of English sentences they would be entirely inappropriate in those circumstances which warranted them if interpreted as utterances of Revolution-Speak. The development of such a language could enable them to cause calculated confusion among the establishment while allowing continuing communication with each other. It might be thought that in this case they ought to be made to pay a price for speaking a foreign language which could be mistaken for English. Even if they never perform any illegalities, the very existence of such a language could be seen as a threat to society and in need of control.

Less extravagantly, the reaction of the linguistic conservative, who objects to the fact that words don't mean what they used to, might also be prompted by the fear of this curious form of subversion. Someone I know particularly objects to the application of the word "minority" to groups, like women, which are actually majorities. What bothers this person is not the identification of women as a group which is discriminated against, but specifically the use of the word "minority" to designate this group. What must motivate this person's objection is the thought that by thus using words, the liberals, with whom this usage originated, are trying to subvert our ability to make unbiased assessments of the issues to which these terms are particularly applicable. The liberals are trying to pull the linguistic rug out from under the conservatives' feet.

Laing might be arguing in a similar vein that the sane condemn someone as mad just because he monkeys about with words like "dead" and "unreal" which play such important parts in the fabric of our moral lives or world-views. It is not what the psychotic says that forces him to pay the price of "being" mad, but how he says it. But although it may be true that ideology can be associated with particular words in a way which makes it a social crime to make free with those words, and although it may well be that the categories of mental illness can be used as measures to repress just such making free, at the end of the day, when all such cases have been discounted, there will surely be many other cases where apparently psychotic people are not merely paying a dreadful price for their choice of words, or even for a genuine challenge to our communal ideology.

I therefore conclude that Laing's approach to understanding psychotics fails because, one might say, there is not enough madness in his method of understanding them. The psychotic becomes the victim of a society which either, completely unaccount-

ably, exacts a price from ontologically insecure foreigners, or slightly more accountably, but with ridiculous xenophobia, punishes people caught using good old words in funny new ways. In either case, this attempt to understand psychotics leaves us unable to say what is wrong with them. Of course, as I said above, Laing does have much to say about what really, and frighteningly, happens in psychosis, but he does not locate the problem at the level where it genuinely seems to exist, namely, at the level of the psychotic's assertion of sentences like "I am dead" and the related sentences and actions that go with them.

II

I turn now to the second approach to the understanding of psychotics. Donald Davidson has argued that in interpretation, and what we are discussing is paradigmatically an issue of interpretation, there is no firm division between how our understanding of the object of interpretation is determined by the meanings we assign to his words, and the beliefs we take him to be expressing.⁶ Thus, in cases where understanding someone on the basis of a given assignment of meanings and beliefs becomes difficult, as it does with the psychotic, we can either reinterpret his words, or reevaluate our assessment of his beliefs. Laing's approach, which we have just discussed, is an example of keeping our assessment of beliefs constant (that is, we don't decide to attribute to the psychotic the belief that he is dead or unreal, in our sense of these words) and of reinterpreting his words. We assign new truth-conditions to his sentences while continuing to interpret him as believing that he is alive in *our* sense of that word, i.e. that he is biologically alive. The approach I am about to consider is the converse of this. Here, the idea is to keep our interpretation of the psychotic's words constant, but to attribute to him radically different beliefs which would rationalize the problematic beliefs in question.

One proponent of this way of understanding psychotics is Silvano Arieti. I quote here a passage by him, concerning a schizophrenic woman who said she was the Virgin Mary.

Her illness made her switch to a primitive form of logic, according to which A becomes B if A has at least one quality of

⁶ See Donald Davidson, *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press, 1984), *passim*.

B. A (in this case the patient Nancy) becomes B (the Virgin Mary) because A has a quality in common with B, namely *being virgin*. Once the patient adopts this type of thinking, her wish is supported by her new logic. She firmly believed she was the Virgin Mary.

This primitive type of logic follows not the classic Aristotelian laws of thought, obeyed automatically by the normal mind, but a principle that was formulated by a psychiatrist named Von Domarus. The principle, in a slightly modified form, says that whereas the normal person accepts identity (that is, he concludes that two objects or two persons are the same) only upon the basis of identical subjects (totality of the subject), the person who uses primitive logic accepts identity based upon identical predicates, that is, a common quality or part. Total sameness is not necessary for identity.

So, when the patient asserts "I am the Virgin Mary", she really is expressing the belief that she is the Virgin Mary, but her beliefs about identity make this a rational belief for her to hold. It may be thought that the question should now simply become whether or not it is rational to hold these other beliefs about identity, but let us give the benefit of the doubt to Arieti and assume that the choice between genuine alternative logics cannot be made on the grounds of rationality. In other words, suppose we have two competing logics, each of which is consistent. Then let us assume for the purposes of present discussion that the choice between them cannot be made on grounds of rationality. So, if Arieti could show that the schizophrenic woman was genuinely using a consistent alternative logic, then we can admit that we could not stigmatize her as irrational merely for her choice of logic (but see the last two paragraphs of this section for a brief discussion of whether Arieti is offering an alternative logic). Even so, I don't think Arieti's suggestion in this case can work.

The axiom, or implicit belief, that is here being attributed to the patient is that for any object x, and for any object y, if there is some property P such that x is P and y is P, then x equals y. Now no-one who is to be comprehensible to us, as Arieti hopes his patient will be, could consistently employ such an axiom with any generality. The patient, for instance, does not identify all yellow objects, or all the objects in the psychiatrist's room, even though there are

⁷ See his *Understanding and Helping the Schizophrenic*, British edition, (Harmondsworth: Penguin Books, 1981), pp. 68-9.

properties which each of the objects in these two sets share. The kinds of difference in judgements which employment of Von Domarus' axiom would entail would obviously be of a much larger scale than Arieti indicates.

It is clear, then, that we do not better understand the psychotic by attributing to him Von Domarus' axiom, since anyone who genuinely held such an axiom would scarcely ever distinguish one thing from another. Such a person would, I venture to say, be totally incomprehensible to us. But Arieti is, presumably, unfamiliar with formal logic and did not intend his remarks to have the conclusion they seem to have. What he meant was, no doubt, that *sometimes* a psychotic patient will identify two objects because they have in common some property, such as being virgin, which is of especial importance in the patient's case, either through the patient's own preoccupation with that property, or perhaps through some significance the property had in the aetiology of the illness. This is no doubt true, but is surely no more than another statement of the problem we faced in the first place. On the supposition that the patient's language is our language, we can see that she identifies herself with the Virgin Mary. What we want to know is how she can do this. Why, given that she is perfectly capable of distinguishing objects in general (i.e. she does not hold Von Domarus' axiom) and given that in general, presumably, she believes in the indiscernability of identicals, does she claim that two objects which have so many different properties, which are so easily discernable, are nonetheless identical? It is precisely because the patient does this that she is considered mad. And what we were attempting to do was to understand her belief. Simply to be told that, on occasion, she has such beliefs is no way of understanding her holding of those beliefs.

Even if we consider the psychical origin or purpose of the belief, and thereby come to understand why, in one sense, the patient has that belief, this will not allow us to characterize the patient as rational in holding that belief. This is because it is part of the concept of belief that beliefs should only represent the world as it really is. Wherever we can see that a belief has arisen because it serves some other purpose, as, for instance, in wishful thinking, we must conclude that a person has that belief irrationally, even if the belief serves some worthwhile function for the person, such as warding off anxiety, and even if the belief happens to be true (which it is not, in the case discussed by Arieti).

Of course, at a higher level, it may be rational to want to have,

and try to bring about, an irrational belief. It is quite rational to try to minimize our suffering, and if we think we can do this effectively by having ourselves hypnotized to believe something we believe with good evidence, before the hypnosis, to be false, then this might be a rational course of action. But there is still a clear sense in which, given that we have good evidence for the falsity of the belief, our holding it is irrational.

Arieti says that “when a statement or a belief that seems absurd to us is expressed with great sincerity and even defended by the patient, it is not done out of stubbornness, caprice, antagonism, obstructiveness or simply because the patient wants to be irrational or difficult. To him, the idea is rational”.⁸ If Arieti was hoping to show that the woman’s idea that she was the Virgin Mary was rational by virtue of her acceptance of some axiom about identity, then he has failed. Attributing to her a belief in Von Domarus’ axiom is of no help since real adherence to such an axiom would lead to total incomprehensibility; yet anything less than this seems to leave the patient as irrational and incomprehensible as we at first thought since we are left with the problem of saying why the patient employs the axiom for some inferences but not for others.

Arieti’s characterization of this woman’s reasoning as employing a different logic is, perhaps, a little misleading. As far as the truth-functional operators go, the woman, even if she did hold Von Domarus’ axiom, does not make use of any inferential practices which are unavailable to classical logic and she does not repudiate any of the inferences of classical logic. Where she does differ from the sane person is over Leibniz’s Law, the indiscernability of identicals. Whereas the sane person holds that if $A = B$, any property of A is a property of B , this woman is prepared to accept cases where $A = B$ (she is the Virgin Mary), but where A and B have different properties – the Virgin Mary lived nearly two thousand years ago, the patient lives now. But whether we count identity as a logical constant, and hence whether this should count as a difference over a logical axiom, or whether it is a difference over an extra-logical axiom which specifies a property of the relation of identity, is not clear.

This confusion is probably not very important in itself, but mentioning it does help to make another point. According to some views of logic the meaning of the logical constants is determined by the axioms which govern them. On such a view, if identity were

⁸ *Ibid.*, p. 68.

considered a logical constant, then in effect, Arieti would have really been taking the same line as Laing. Even if identity were not considered a logical constant, Arieti would still have done the same as Laing so long as one takes the axiom, either Von Domarus' or Leibniz's, as determining its meaning. In such a case, Arieti's claim would be that, when the psychotic utters the sentence "I am identical to the Virgin Mary", we, the sane, should not describe her utterance as expressing the belief that the speaker is identical to the Virgin Mary. For us, "to be identical to" means to share all properties whereas for the psychotic, "to be identical to" means to share some property. Hence, we should understand the psychotic's sentence as expressing her belief that she shares some, but not every, property with the Virgin Mary – and we should hardly take her to be insane for having that belief. This is just the sort of conclusion that followed from Laing's views.

III

It is time to say what I think regarding the understanding of madness. What I want to suggest, and rather briefly argue for, is that we should characterize madness precisely by the difficulties it presents for understanding. It is just what makes the psychotic mad that we should fail, or be frustrated, in our attempts to understand him. But this failure is not due to some weakness in our ability to understand or, correlatively, to the bizarreness of what the psychotic says or believes. It is due to a real challenge that understanding in general is liable to face in certain extreme circumstances, stemming from the very nature of what it is that is understood, i.e. mental content and action. Let me now elaborate on the theory of mind of which this threat to understanding is a consequence.

The kind of theory which is at issue is a theory which says which kinds of facts constitute the propositional contents of mental states. Theories which address this question may be divided into two sorts, which I will call "realist" and "idealist" (though I wouldn't like these labels to interfere with an understanding of what they label). Realist theories of content are those which hold that the content of a given mental state is constituted by facts which are intrinsic to that state. This means that they will not see a connection between the constitution of the contents of some person's mental states and the connections between mental states which obtain in virtue of their contents. Since the relations which

hold between mental states in virtue of their contents will be such relations as entailing, being a good reason for, and so on, realist theories see no connection between the constitution of the contents of mental states and rationality. On the realist view, there would be nothing odd (logically odd, that is) about a creature having a set of mental states which were either not rationally or even irrationally connected.

Fodor, an exponent of a realist view, admits as much when he says that what makes it true that someone's propositional attitudes cause actions they rationalize *in virtue of their content*, i.e., what does account for the rational connections between reasons and actions (and by implication, between mental states and other mental states), is that "there are true, contingent counterfactuals which relate mental state *tokens* in virtue of their content" and that these counterfactuals hold because "there are true contingent generalizations which relate mental state *types* in virtue of their contents".⁹ That these connections are contingent means that there is no contradiction in supposing that they fail to hold and that a person's mental states cause actions that they do not rationalize in virtue of their contents.

Not surprisingly, such realist views will have no problem with, and hence nothing special to say about, cases of irrationality. A psychotic could have just the beliefs he appears to have, when he asserts the sentence "I am dead", because what makes some state of his brain a belief that he is dead is completely independent of facts about whether it is rational for him to have that belief. All that one can say about such a person is that his "belief-mechanism" has gone wrong for whatever reasons, psychological or physiological.

In the rest of this paper, I want to see what idealist theories have to say about irrationality and madness. I will offer no arguments here as to why one should prefer an idealist to a realist theory of content (though I think one should). This is a hotly debated question in philosophy and in this context I could not do justice to the issue. But I do think it is a point, however small, in favour of idealist theories that they should allow us to take just the view they do of madness.¹⁰ As will become apparent, I think such a view is simply not open to proponents of a realist theory of content.

For what I am calling idealist theories of content, theories

⁹ Fodor, "Propositional Attitudes", in *Representations* (Brighton: Harvester Press, 1981), p. 183.

¹⁰ See my "Freud's Ambiguous Concepts", in *Journal of Speculative Philosophy* vol. 3, no. 2 (1989) for another similar example of how idealist theories can do something interesting that realist theories cannot.

advocated by, for instance, Davidson and Dennett, there is no divorce of content and rationality. What constitutes a person's having a particular belief is not independent of how that belief connects with perception, action and other mental states. Philosophers who have held such theories have spent a lot of energy in specifying what the constraints are which must be met for organisms to have various types of mental states. Whatever these constraints are, once we understand them, and of course it is part of the claim of these philosophers that understanding them requires nothing abstruse or esoteric, we can use them to interpret organisms. If the principles which govern our interpretation mirror the constraints which must hold between mental states, then interpretation is not merely a means of finding out what somebody believes and desires, but an explicit articulation of the very things that constitute those beliefs and desires.

What guide us in interpretation, and what form the constraints on how mental states must be related if they are to exist at all, are the normative principles of rationality to which Davidson gives the collective name "the Principle of Charity".¹¹ These principles require that, for example, beliefs are not false (by our standards, of course, but then what other standards could we use); that the descriptions given to actions under which they are intentional are such as to conform to the agent's best judgement about what he ought to do; that patterns of preference be transitive and non-symmetric, and so on.

Unless we interpret someone as conforming to these principles, we cannot, on an idealist theory, interpret him at all. Since interpretation is an articulation of what constitutes content, to say that someone is rationally uninterpretable is equivalent to saying that there is no mental content there to be discovered. The constraints on interpretation implicitly define the concepts, such as belief, desire and action, that they govern. Thus, as Davidson stresses time and time again, "whether people might not actually *be* approximately rational and consistent in their patterns of belief and desire . . . cannot be a factual question: if a creature has propositional attitudes then that creature is approximately rational".¹² This is just the kind of connection between content and rationality that realist theories seek to deny.

¹¹ That this is a collective name for a group of principles, and not the name of a particular one of them, is a matter which might require some exegetical support – support which, I believe, it could be given, but which is inappropriate in this context.

¹² Davidson, "Reply to J. J. C. Smart", in Bruce Vermazen and Merrill Hintikka (eds.), *Essays on Davidson. Actions and Events* (Oxford: Clarendon Press, 1985), p. 245.

With this idealist, as opposed to the realist, approach there is a problem with irrationality. Cases of apparent irrationality are cases where a person's mental states, or actions, seem to be at odds with the normative constraints. As interpreters, we thus face a dilemma. If we accept the constraints at face value, we can attribute various (perhaps unconscious) mental states to rationalize the apparent irrationality. In that case, we explain it away. On the other hand, if we accept the irrationality at face value, we countenance an exception to the normative constraints. Yet it is these normative constraints which constitute mental content, and their violation, therefore, threatens to undermine the basis of rationality and irrationality alike.¹³

The question naturally arises of how much irrationality we can attribute without undermining the possibility of interpretation. It seems clear that we should allow a certain amount of irrationality before we have to abandon the project of seeing someone as having beliefs and desires. But where, exactly, should we draw the limit?

I don't think there is, or need be, any way of giving a precise answer to this. One of the consequences of idealist theories of content is the kind of indeterminacy of content we have come to associate with Quine and Davidson. To expect a definite answer to the question of how much irrationality we can tolerate is to rely, implicitly, on a realist conception of content according to which there are facts of the matter about people's beliefs and other mental states which hold independently of the application of the necessarily somewhat flexible rationality constraints. Wanting to know exactly how flexible these constraints are primarily arises from thinking that there is some underlying mental reality to which they must conform.

The point of the idealist theories is that there is no such underlying mental reality. The Principle of Charity is not an epistemological constraint but a constitutive one. Instead of thinking that since there are determinate facts of the matter about people's mental states there must be a determinable limit to how flexible the rationality constraints are, we should, according to the idealist theories, accept that the constraints are flexible and hence that there is no independent, underlying mental reality to which they must be faithful.

¹³ Davidson, perhaps more than any other philosopher, has seen the problem posed by irrationality, and vigorously attempted to solve it. See my forthcoming book on Davidson (Oxford: Polity Press) for a discussion of why I think these attempts to deal with irrationality are unsuccessful. What is important to me in this paper is the existence of the problem, rather than the attempt to solve it.

This brings us round to the question of understanding psychosis. The psychotic is, apparently, extremely irrational. Both the attempts at understanding madness that we looked at recognized some connection between understanding and rationality. Each of them, in trying to show how psychotics can be understood, tried to show that in one way or another, the subjects were rational, one of them by re-interpreting the subjects' language, the other by attributing to them strange beliefs about identity. I have tried to say what I think in particular is wrong with these attempts, but ultimately they seem wrong because they ignore our basic intuition that madness *is* an extreme form of irrationality. What makes it so strange for someone to say "I am dead" or "I am the Virgin Mary" is that these express irrational beliefs. Any theory which ends up showing that these are not expressions of irrational beliefs, either by showing that they express different beliefs (i.e. have different meanings) or by attributing to the psychotic unusual beliefs about identity, will end up by showing that such people are, after all, not mad. Of course, it would be wonderful if no-one were mad, but as long as we think there is something to the idea of madness, we had better give up the idea that such people are not irrational.

But if they are irrational, then, on the idealist theory of content sketched above, this very fact means that they will be difficult to understand. And the more irrational they are, the more mad they are, the more difficult they will be to comprehend. The more we try and accept them at face value, the harder it is to continue to apply to them the concepts of belief, desire, and action. Even everyday cases of irrationality create a problem for interpretation, and hence are a challenge (albeit usually an insignificant one) to applying the concepts of belief, desire, and action to the irrational person. If madness involves extreme irrationality, the challenge is far more serious. At the limiting case, with the (let us hope caricatured) picture of the raving lunatic, we will have to say that such a person can no longer really be thought of as having any beliefs or desires, or as performing any intentional actions, at all. Those constraints which implicitly define these concepts, the Principle of Charity, cannot be satisfied. We simply have a biological mechanism gone haywire.

If, at the limiting case, we have just a biological mechanism gone haywire, then it might seem as if we are close to the realist theories again. For did I not say, in describing them, that they would see madness simply as a (biological or psychological) failure of an organism to have its mental states connected to each other and to

actions in rational ways? So, for such a theory, would not the psychotic also be a biological mechanism gone haywire?

In fact, the similarity between the realist and idealist pictures here is only superficial. On the idealist theories, at this extreme position, madness itself disappears. The concepts which the realist uses to characterize madness are those whose application in the extreme, for the idealist, coincides with the disappearance of the very phenomenon the realist takes them to characterize. Madness arises as the conflict between the demands of rational interpretation and the difficulty in achieving it. It is only when we are committed enough to interpretation to be frustrated by our failure to interpret sufficiently that we judge someone mad. If the mental states of a person are thought of as a pattern, one could say that one can only see the breakdown in this pattern when one can still discern something of the pattern itself. Concepts like the breakdown of a pattern, or like madness, only apply when the very things of which they announce the destruction are to some extent still present.

It may be objected that I have only succeeded in saying something about how we apply the concept of madness, and not about what madness is. It is true that I have been largely concerned, in the last few paragraphs, with the application of the concept, but I do not think that this justifies the claim that I have ceased discussing the phenomenon of madness. Remember that, for the idealist theory of content on which my discussion is predicated, the constraints on interpretation mirror the facts which determine, or constitute, propositional content. Thus, talking about the way in which we attribute beliefs and desires can also be seen as talk of what belief and desire actually are. It is this which makes such theories idealist. Madness is being viewed here as a concept which goes with the concepts of propositional attitude psychology. Hence, for madness as well, there is no gap between what the phenomenon is and what constrains how we apply the concept. I fully accept that some argument is needed to justify this position in general, but I have already explained that this is not something I propose to do in this paper: the task is far too large.¹⁴ All I am concerned with here is to show what consequences the idealist theory has for the question of understanding madness.

The crucial point in the view of madness being broached is the conflict between two sets of needs, the need to find a large amount of reason if we are to find any mental content at all (and hence any

¹⁴ I attempt some such argument in my book on Davidson (see previous note for details).

mental content which can be held irrationally), and the need to do justice to the genuine and dramatic irrationality of the insane. This conflict, between the needs of interpretation and the difficulty in meeting those needs, is a conflict which arises only out of the particular idealist view of rationality and mental content which I have attempted to describe above. It will simply not arise for a realist theory which, like Fodor's, does not make the principles of rationality themselves enter into the constitution of the contents of actual mental states.

The conflict is, of course, just the problem that the idealist approach has over irrationality in general. As Davidson says: "the underlying paradox of irrationality, from which no theory can entirely escape, is this: if we explain it too well, we turn it into a concealed form of rationality; while if we assign incoherence too glibly, we merely compromise our ability to diagnose irrationality by withdrawing the background of rationality needed to justify any diagnosis at all".¹⁵ What I want to emphasize is that in a sense, this problem is actually a virtue. Madness, the disarray of reason, is, if one thinks about it, an exceedingly puzzling phenomenon. That a theory of rationality has difficulties over irrationality reflects the fact that irrationality and madness are themselves undeniable problems for reason. For any theory which doesn't have a problem with madness, such as a realist theory which dissociates rationality from mental content, madness becomes nothing more than a phenomenon outside of reason; reason itself would not be disturbed by madness. Such a view, to use Foucault's words, "thrusts into oblivion all those stammered, imperfect words without fixed syntax in which the exchange between madness and reason [is] made".¹⁶

By contrast, on the idealist view, madness is not exiled from reason. We understand it better by seeing precisely that it is a problem for understanding that lies barely concealed within reason itself. Lacan expresses this when he writes: "Not only can man's being not be understood without madness, it would not be man's being if it did not bear madness within itself as the limit of his freedom".¹⁷ The idealist approach to mental content shows that rationality is a continuous struggle, the struggle of interpretation. The attribution of mental content is always being measured against an ideal of rationality which can never be too exacting. Conse-

¹⁵ "Paradoxes of Irrationality", in Richard Wollheim and James Hopkins (eds.), *Philosophical Essays on Freud* (Cambridge: Cambridge University Press, 1982), p. 303.

¹⁶ *Madness and Civilization* (London: Tavistock Publications, 1967), p. xii.

¹⁷ *Ecrits: A Selection* (London: Tavistock Publications, 1977), p. 215.

quently, madness is not just the Dionysian other to the Apollonian coolness of reason. It is not a phenomenon which lies outside of reason, leaving reason unaffected. Madness points to the perpetual instability of reason and our very difficulty in comprehending the insane reminds us of the continuous effort required to understand the sane.¹⁸

*26 Deloraine House
Tanners Hill
London SE8 4PY*

¹⁸ I am grateful to Jo Wolff for making very helpful comments on not one, but two earlier drafts of this paper.